Comparative Study of Artificial Neural Network and Response Surface Methodology for Modelling and Optimization the Adsorption Capacity of Fluoride onto Apatitic Tricalcium Phosphate

M.Mourabet*, A. El Rhilassi, M.Bennani-Ziatni, A. Taitai

Team Chemistry and Valorization of Inorganic Phosphates, Department of Chemistry, Faculty of Sciences, BP 133, Kenitra, Morocco

*Corresponding Author: mrrab.et@hotmail.com

Abstract In this study, Response Surface Methodology (RSM) and Artificial Neural Network (ANN) were employed to develop an approach for the evaluation of fluoride adsorption process. A batch adsorption process was performed usingapatitic tricalcium phosphate as an adsorbent, to remove fluoride ions from aqueous solutions. The effects of process variables which are pH, adsorbent mass, initial concentration, and temperature, on the adsorption capacity (\(\text{q}_e\) (mg/g)) of fluoride were investigated through three-levels, four-factors Box-Behnken (BBD) designs. Same design was also utilized to obtain a training set for ANN. The results of two methodologies were compared for their predictive capabilities in terms of the coefficient of determination \(R^2\), root mean square error \(RMSE\), and the absolute average deviation \(AAD\) based on the experimental data set. The results showed that the ANN model is much more accurate in prediction as compared to BBD.

Keywords Response Surface Methodology, Box-Behnken Designs, Artificial Neural Network, Adsorption Capacity, Fluoride

1. Introduction

Response Surface methodology (RSM), introduced by Box and Willson [1], is a collection of mathematical and statistical technique useful for analyzing problems in which several independent variables influence a dependent variable or response and the goal is to optimize the response[2]. The design procedure of response surface methodology is as follows [3]:

1. Designing of a series of experiments for adequate and reliable measurement of the response of interest.
2. Developing a mathematical model of the second order response surface with the best fittings.
3. Finding the optimal set of experimental variables that produce a maximum or minimum value of response.
4. Representing the direct and interactive effects of process variables through two and three dimensional plots.

A second-order (quadratic) model is commonly used in response surface methodology:

\[
y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + \epsilon \tag{1}
\]

where \(y\) is the process response or output (dependent variable); \(x_1, x_2, ..., x_k\) are the input factors; \(\beta_0, \beta_{ii} (i = 1,2,...,k), \beta_{ij} (i = 1,2,...,k; j = 1,2,...,k)\) are unknown parameters and \(\epsilon\) is a random experimental error term assumed to have a zero mean.

The most common second-order RSM designs are Box–Behnken (BBD) design and the Central Composite Design (CCD) with its variants. Box–Behnken design is a three-level design based on the construction of a balanced incomplete block design [4]. The BBD is an efficient option for fitting response surfaces using three evenly spaced levels [5]. Fig.1 (a) illustrates a Box–Behnken design for three factors. The number of experimental points \(N\) is defined by the expression \(N = 2k(k-1) + c_p\), where \(k\) is the number of variables and \(c_p\) is the number of center points [5].

Artificial neural network (ANN) is a computational model that tries to simulate the structure and functionalities of biological neural networks. It is a parallel distribution processing system composed of nodes (neurons) and connections (weights), and is based on the principle that a highly interconnected system of simple processing elements that can learn complex interrelationships between independent and dependent variables [6]. A typical neural network has an input layer, one or more hidden layer, and an
output layer. The number of input and output neurons is fixed by the nature of the problem.

The objective of a neural network is to compute output values from input values by some internal calculation [7]. Among the various types of neural networks, the feed-forward network with back propagation algorithm is the most widely used in different applications [8-13]. In the feed-forward ANN, the information is processed in the forward direction from input layer to hidden and then output layer obtained as the output of the network. The function of the hidden neurons is to intervene between the external input and the network response. The connection between inputs, hidden and output layers consist of weights and biases that are considered parameters of the neural network.

### Figure 1

(a) Box-Behnken design for three factors. (b) Structure of a three-layer feed-forward network of $m$ inputs, hidden layer with $k$ neurons, output layer and 1 output.

For a three-layer feed-forward ANN (Fig.1 (b)), the $\ell$th estimated response, $y^\ell$, that can be expressed:

$$y^\ell = \psi_0 \left( b^\ell + \sum_{j=1}^{k} w_{ij} \psi_h \left( \sum_{i=1}^{m} w_{ij} x_i^\ell + b_j \right) \right)$$  \hspace{1cm} (2)

where, $W = (w_{ij})$ and $B = (b_j)$ are the weight matrix and bias matrix of hidden layer, $W' = (w_1', w_2', ..., w_k')$ and $b'$ (scalar) are the weight matrix and bias of output layer, $Y = (y^1, y^2, ..., y^\ell)'$ is response vector, $X_i$ is the $i$th input vector ($X_i = (x_1^i, x_2^i, ..., x_m^i)$; $i = 1, 2, ..., m$) and $\psi_0$, $\psi_h$ are transfer functions of hidden layer and output layer respectively.

Three types of commonly used transfer functions are as follows:

- **Linear transfer function**
  $$\psi(x) = x, \quad -\infty < \psi(x) < +\infty$$  \hspace{1cm} (3)

- **Sigmoid transfer function**
  $$\psi(x) = \frac{1}{1+e^{-x}}, \quad 0 \leq \psi(x) \leq 1$$  \hspace{1cm} (4)

- **Hyperbolic tangent transfer function**
  $$\psi(x) = \frac{2}{1+e^{-2x}} - 1, \quad -1 \leq \psi(x) \leq 1$$  \hspace{1cm} (5)

Response surface methodology (RSM) and artificial neural network (ANN) methods are now successfully used jointly for both modeling and optimization purposes in many areas of science and engineering [13, 14-20]. In these studies, ANN and RSM techniques were compared for their predictive and generalization capabilities, sensitivity analysis and optimization abilities. It was found that the ANN model fit the data better and has higher predictive capability than RSM, even with the limited number of experiments.

However, there are only a few applications of ANN with RSM for the modeling of adsorption process [21-23]. Hence, the main motivation behind this study is the using the RSM and ANN methodologies for predicting the adsorption capacity of fluoride by apatitic tricalcium phosphate and the results which were obtained through RSM were then compared with those through ANN.

### 2. Material and Methods

#### 2.1. Adsorbent

The apatitic tricalcium phosphate ($Ca_9$ (HPO$_4$) (PO$_4$)$_5$(OH)) powders were prepared by an aqueous double decomposition of the salts of calcium and of phosphate [24-25].

#### 2.2. Batch Adsorption Experiments

Experiments were carried out in 125mL glass bottle by adding the desired amount of apatitic tricalcium phosphate in 50 mL of fluoride solutions of desired initial concentration, pH, and temperature. The mixture was agitated at fixed contact time (90min) and at 300 rpm. The solution was then filtered and the residual fluoride ion concentration analyzed electrochemically with a fluoride ion-selective electrode (Orion, USA) by the use of total ionic strength adjustment buffer (TISAB) solution. The adsorption capacity $q_e$ (mg/g) of fluoride adsorbed was evaluated by following expression:

$$q_e = \frac{(C_0-C_e)\nu}{m} \hspace{1cm} (6)$$

where, $C_0$, initial concentration of fluoride (mg/L), $C_e$, concentration of fluoride in solution at equilibrium time, $\nu$, solution volume (L), and $m$, adsorbent dosage (g).
2.3. Experimental Design

2.3.1. Response Surface Methodology

A three-level four factor Box-Behnken experimental design was generated with the Design- Expert 7.0.0 software. Initial concentration (30–60mg/L), pH (4–11), adsorbent dose (0.1–0.3g) and temperature (20–40°C) were input variables, the factor levels were coded as −1 (low), 0 (central point), and 1 (high). The design of real experiments is given in Table 1.

The response variable, \( q_e \) (adsorption capacity, mg/g) can be expressed as a function of the independent process variables (initial concentration \( x_1 \), pH \( x_2 \), adsorbent dose \( x_3 \) and temperature \( x_4 \)) according to the following response surface quadratic model:

\[
q_e = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{12} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{44} x_4^2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 + \beta_{34} x_3 x_4
\]

(7)

The coefficients, i.e. the model constant \( \beta_0 \), the linear terms (\( \beta_1, \beta_2, \beta_3, \beta_4 \)), the quadratic terms (\( \beta_{12}, \beta_{22}, \beta_{33}, \beta_{44} \)) and the interaction terms (\( \beta_{12}, \beta_{13}, \beta_{14}, \beta_{23}, \beta_{24}, \beta_{34} \)), have been estimated from the experimental results applying least square method.

2.3.2. Artificial Neural Network

In this study, Neural Network Toolbox 8.0 of MATLAB mathematical software was used for simulation. The same experimental data, which had been used for the RSM design, were also employed in designing the artificial neural network. The input variables were initial concentration, pH, adsorbent dose and temperature. The corresponding adsorption capacity was used as a target.

The data were randomly divided into three groups, 70% in the training set, 15% in the validation set and 15% in the test set.

The tangent sigmoid transfer function (tansig) at hidden layer and a linear transfer function (purelin) at output layer were used. The similar transfer function was also used by Oguz and Ersol [26], Elmolla et al [27] and Khajeh et al [11]. The training function selected for the network is ‘trainlm’. ‘Trainlm’ is a network training function that updates weight and bias values according to the Lavenberg-Marquardt algorithm.

All variables and response were normalized between 0 and 1 for the reduction of network error and higher homogeneous results. The normalization equation applied is as follows:

\[
y_n = \frac{y_a - y_{\text{min}}}{y_{\text{max}} - y_{\text{min}}} \quad (8)
\]

where \( y_n \), \( y_a \), \( y_{\text{min}} \) and \( y_{\text{max}} \) are normalized value, actual value, minimum value, and maximum value, respectively.

3. Results and Discussion

3.1. Response Surface Methodology

Based on the experimental results of BBD in Table 1, a quadratic polynomial was established to identify the relationship between adsorption capacity and process variables. The resulting RSM model equation is following:

\[
d_e = 8.8117 - 0.0406 x_1 - 0.3381 x_2 - 14.0445 x_3 + 0.0582 x_4 + 0.0012 x_1^2 - 0.0084 x_2^2 - 37.3792 x_3^2 + 0.0118 x_1 x_2 + 0.0481 x_2 x_3 - 0.1915 x_1 x_3 + 0.0003 x_1 x_4 + 2.8107 x_2 x_3 - 0.0167 x_2 x_4 + 0.0645 x_3 x_4 \quad (9)
\]

From Eq. (9), it can be seen that the initial concentration, pH, and adsorbent dose, have negative effect on the adsorption capacity compared to the temperature, which has a positive effect on the adsorption capacity. The experimental results and the predicted values obtained from the model (Eq. (9)) were compared. According Fig.2, it was found that the predicted values matched the experimental values reasonably well with \( R^2 = 0.927 \). This implies that 92.7% of the variations for adsorption capacity are explained by the independent variables, and this also means that the model does not explain only about 7.3% of variation. In addition, the value of adjusted determination coefficient \( R^2_a = 0.853 \) was also high, showing a significance of the model.

Table 2 gives the results of the quadratic response surface
model fitting in the form of analysis of variance (ANOVA). The analysis of variance is essential to test significance and adequacy of the model. It subdivides the total variation of the results in two sources of variation, the model and the experimental error, shows whether the variation from the model is significant when compared to the variation due to residual error [28]. The Fisher’s F-test value, which is the ratio between the mean square of the model and the residual error, performs this comparison [29, 30]. If the model is a good predictor of the experimental results, F-value should be greater than the tabulated value of F-distribution for a certain number of degrees of freedom in the model at a level of significance. The F-value obtained, 12.69, is greater than the F value (2.47 at 95% significance) obtained from the standard distribution table, confirming the adequacy of the model fits. In addition, the p-value was found to be < 0.0001, which indicated that the model was highly statistically significant.

The “Lack of Fit Test” compares the residual error to the pure error from replicated design points. The “Lack of Fit F-value” of 1.09 implies the Lack of Fit is not significant relative to the pure error. There is a 51.07% chance of occurrence of noise, indicating significance of the model. The significance of each term was determined by p-value (Prob>F), which is listed in Table 2.

As seen in this table that the terms \( x_1 \), \( x_2 \), \( x_3 \), \( x_4 \) and \( x_2x_3 \) were significant, with very small p-values (p < 0.05). The other term coefficients were not significant (p > 0.05).

The Pareto analysis [31] was carried out to check the percentage effect of each factor. In fact, this analysis calculates the percentage effect (\( P_i \)) of each factor on the response, according to the following relation:

\[
P_i = \left( \frac{\beta_i^2}{\sum \beta_i^2} \right) \times 100 \quad (i \neq 0)
\]

where \( \beta_i \) is the regression coefficient of individual process variable.

Fig. 3 shows the Pareto graphic analysis. As can be seen in this figure, among the variables, the adsorbent dose (\( \beta_3 \), 12.308% and \( \beta_3^2 \), 87.188%) produce the main effect on the adsorption capacity.

### 3.2. Artificial Neural Network

In order to determine the optimum number of neurons in the hidden layer, a series of topologies was examined. The mean square error (MSE) was used as the error function. It measures the performance of the network. Moreover, the correlation coefficient (\( r \)) was used as a measure of the predictive ability of the network. ANN optimization process required network training to minimize the error function (MSE) by searching for a set of connection weights that can enable the ANN to produce outputs that are identical or possibly equal to target values.

After repeated trials, it was found that a network with 11 hidden neurons produced the best performance. The optimal architecture of ANN model in this case is shown in Fig. 4. It
has three-layer ANN, with tangent sigmoid transfer function (tansig) at hidden layer with 11 neurons and linear transfer function (purelin) at output layer. The MSE value was found to be 0.0003 (Fig. 5). A regression analysis between ANN outputs and the experimental data was carried out. This ANN model indicated a precise and effective prediction of the experimental data with a correlation coefficient ($r$) of 0.989, 0.971, 0.996 and 0.989 for training, validation, testing and all data, respectively (Fig. 6).

The weights and biases associated with the final trained network are given in Table 3. The ANN model can be presented mathematically as an input–output composite mapping:

$$Y_n(X) = \phi^{(2)}(\mathcal{LW}[2,1] \phi^{(1)}(\mathcal{IW}[1,1]X + b[1]) + b[2])$$

(11)

where $Y_n$ denotes the vector of the normalized output (network predictions), $X$ is the vector of the input variables, $\phi^{(1)}$ is the vector of tansig transfer function corresponding to the hidden layer, $\phi^{(2)}$ is the vector of purelin transfer function corresponding to the output layer, $\mathcal{IW}[1,1]$ is the input weight matrix, $\mathcal{LW}[2,1]$ is the layer weight vector, $b[1]$ is the bias vector, and $b[2]$ is the bias scalar.

Table 3. Optimal values of the network weights and biases.
To evaluate the relative importance of the input variables on the Output variable ($q_e$ (mg/g)), neural net weight matrix and Garson equation were used in the evaluation processes. Garson proposed the equation based on the partitioning of connection weights as shown in Eq. (12) [27, 32-33]:

$$J_j = \frac{\sum_{m=1}^{N_o} \left[ \sum_{i=1}^{N_i} \frac{w_{m,i}^j}{\sum_{k=1}^{N_h} w_{m,k}^j} \right] w_{m,j}^o}{\sum_{j=1}^{N_o} \sum_{m=1}^{N_o} \left[ \sum_{i=1}^{N_i} \frac{w_{m,i}^j}{\sum_{k=1}^{N_h} w_{m,k}^j} \right] w_{m,j}^o}$$ (12)

where, $J_j$ is the relative significance of the $j$th input variable on the output variable, $N_i$ and $N_h$ are the number of input and hidden neurons, respectively. $W$ is connection weight, the superscripts “i”, “h”, and “O” represents the input, hidden, and output layers, respectively, while the subscripts “h”, “m”, and “n” refer to input, hidden, and output neurons, respectively.

The relative significance (Table 4) of the four input variables computed by the Garson equation showed that all variables have strong effect on adsorption capacity ($q_e$). Adsorbent dose and pH are the most significant variables with 38.24 and 26.11%, respectively, followed by the initial concentration of fluoride (19.49%), and finally the temperature (16.14%).

<table>
<thead>
<tr>
<th>Input Variable</th>
<th>Importance(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Concentration</td>
<td>19.49</td>
</tr>
<tr>
<td>pH</td>
<td>26.11</td>
</tr>
<tr>
<td>Adsorbent dose</td>
<td>38.24</td>
</tr>
<tr>
<td>Temperature</td>
<td>16.14</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
</tr>
</tbody>
</table>

### 3.3. Comparison of ANN and RSM Models

The comparison of RSM and ANN methodologies for predicted experimental results was done in terms of coefficient of determination ($R^2$), root mean squared error (RMSE) and average absolute deviation (AAD) which can be defined as follows:

$$RMSE = \left( \frac{1}{n} \sum_{i=1}^{n} (y_{i,exp} - y_{i,pred})^2 \right)^{1/2}$$ (14)

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_{i,exp} - y_{i,pred})^2}{\sum_{i=1}^{n} (y_{i,exp} - \bar{y}_{exp})^2}$$ (15)

$$AAD = \left\{ \sum_{i=1}^{n} \left[ \frac{|y_{i,pred} - y_{i,exp}|}{y_{i,exp}} \right] \right\} \times 100$$ (16)

where $y_{i,pred}$ was the predicted value by ANN model, $y_{i,exp}$ was the experimental value, $n$ was the number of data, and $\bar{y}_{m}$ was the average of the experimental value.

The comparative values RMSE, AAD, and $R^2$ are given in Table 5. The root mean squared error (RMSE) for the design matrix by RSM and ANN is 0.0942 and 0.0262, the coefficient of determination ($R^2$) is 0.927 and 0.979, and the absolute average deviation (AAD) is 5.110 and 1.320. The deviation of predicted response from experimental data (table1) for both methodologies is shown in Fig. 7. These results indicate that the ANN model showed a clear superiority over RSM for both data fitting and estimation capabilities. This higher predictive accuracy of ANN can be attributed to its universal ability to approximate nonlinearity of the system whereas RSM is only restricted to a second-order polynomial [34, 35]. In addition, it is to be noted that though RSM has the advantage of giving a regression equation for prediction and showing the effect of experimental factors and their interactions on response in comparison with ANN, ANN does not require a standard experimental design to build the model [23]. Another advantage of ANN model is flexible and permits to add new experimental data to build a trustable ANN model. In contrast, ANN methodology may require a greater number of experiments than RSM [23].

### Table 5. Comparison of RSM and ANN.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Box-Behnken</th>
<th>Design data</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>0.0942</td>
<td>0.0262</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.927</td>
<td>0.979</td>
</tr>
<tr>
<td>AAD</td>
<td>5.110</td>
<td>1.320</td>
</tr>
</tbody>
</table>

### Figure 7. The scatter plot of RSM and ANN model predicted values versus actual values for Box-Behnken design matrix.

### 4. Conclusion

In this study, the effects of pH, adsorbent mass, initial concentration, and temperature, on the adsorption capacity of fluoride onto apatitic tricalcium phosphate were investigated using RSM and ANN methods. The root mean square error (RMSE), coefficient of determination ($R^2$) and absolute average deviation (AAD) were used together to compare the performance of the RSM and ANN models. The ANN model was found to have higher predictive capability than RSM model even with limited number of experiments.
Thus, it can be concluded that even though RSM is most widely used method for adsorption optimization, the ANN methodology may present a better alternative.

REFERENCES


